

Metabolic networks are almost nonfractal: A comprehensive evaluation

Kazuhiro Takemoto^{1,*}

¹ *Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, Iizuka, Fukuoka 820-8502, Japan*
(Dated: July 31, 2014)

Network self-similarity or fractality are widely accepted as an important topological property of metabolic networks; however, recent studies cast doubt on the reality of self-similarity in the networks. Therefore, we perform a comprehensive evaluation of metabolic network fractality using a box-covering method with an earlier version and the latest version of metabolic networks, and demonstrate that the latest metabolic networks are almost self-dissimilar, while the earlier ones are fractal, as reported in a number of previous studies. This result may be because the networks were randomized because of an increase in network density due to database updates, suggesting that the previously observed network fractality was due to a lack of available data on metabolic reactions. This finding may not entirely discount the importance of self-similarity of metabolic networks. Rather, it highlights the need for a more suitable definition of network fractality and a more careful examination of self-similarity of metabolic networks.

PACS numbers: 89.75.Hc, 61.43.-j

I. INTRODUCTION

Metabolism can be defined as a series of biochemical reactions, and it is often represented as a network. Generically, when metabolic networks are depicted in a schematic fashion, metabolites are portrayed as nodes and reactions as edges [1–3]. In recent years, several new technologies and high-throughput methods have generated a massive quantity of genomic and metabolic network data; thus, the overall picture of the metabolic world has gradually become clearer (reviewed in [4–7], for example).

In particular, several studies, until now, have demonstrated a scale-free property (i.e., a power-law distribution of the number of links per node or degree) in metabolic networks as well as in other types of networks such as social networks and World Wide Web [8].

The scale-free property (power-law degree distribution) is often used in the sense of scale invariance (i.e., self-similarity). In addition to this, a hierarchical organization (i.e., self-similar nesting of different modules) in complex networks [9], including metabolic networks [10], also implies network self-similarity.

Song et al. [11] have shown that a number of real-world networks, including metabolic networks, have self-similarity (fractality) using a box-counting method for complex networks, in which a box is defined as a set of nodes between which distances (shortest path lengths) are less than a box size (l_B). In particular, a fractal network shows a power-law relationship in the minimum number of boxes ($N_B(l_B)$) covering nodes with linear size l_B :

$$N_B(l_B) \propto l_B^{-d_B},$$

where d_B is a fractal dimension.

According to this previous study, in this paper, network *self-similarity* is defined by $N_B(l_B)$ (see Sec. II C for details), and it is synonymous with network *fractality*. However, note that the scale-free property (power-law degree distribution) does not necessarily correspond to network fractality. In particular, Song et al. [11] have demonstrated that scale-free random networks (i.e., Barabási–Albert (BA) model [8]) show no fractality while they have power-law degree distributions; in particular, $N_B(l_B)$ of BA model networks exponentially decays with l_B .

Network fractality originates from a negative degree–degree correlation (i.e., disassortativity) [12] or the repulsion between hubs [13], and it helps not only to detect functional modules in biological networks but also to understand network evolution [13] and the utility and the redundancy in networked systems [14]. Therefore, a number of investigators have accepted self-similarity of metabolic networks, and they have used metabolic networks as an example of fractal networks [11–17].

However, network self-similarity has sometimes been criticized [18]. Metabolic networks are suggested to be self-dissimilar and scale-rich [19, 20] because they are compartmentalized [21] and the degree distributions among different biosynthetic modules are very different, although the definition of modules (boxes) is different between these studies and previous studies (e.g., [11, 13]).

This inconsistent conclusion emphasizes the need for a more careful examination of network fractality. In particular, previous studies (e.g., [11, 13]) have the following problems: consideration of only a few representative organisms and the use of an old dataset of metabolic networks published in 2000 [1]. Thus, metabolic network fractality is still debatable. In particular, universality of metabolic network fractality is still poorly evaluated. Moreover, metabolic network analysis strongly depends on the accuracy of metabolic information. An opposite conclusion may be derived from a comparison between an earlier version and the latest version of metabolic net-

*Electronic address: takemoto@bio.kyutech.ac.jp

works [22].

The latest data regarding metabolic information are collected in several databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) [23] and the Encyclopedia of Metabolic Pathways (MetaCyc) [24]; these databases are widely used and contain the metabolic pathways of many living organisms.

In this study, therefore, we perform a comprehensive evaluation of metabolic network fractality using a box-covering method with an earlier version and the latest version of metabolic networks of a large number of organisms, and we demonstrate that the latest version of metabolic networks show self-dissimilarity while the earlier version of metabolic networks are fractal, as reported in [11, 13]. Furthermore, we discuss a possible origin of the vanishment of network fractality due to database updates, importance of a careful examination of biological networks, more suitable definition of network fractality.

II. MATERIALS AND METHODS

A. Selection of organisms

We used previously published lists of prokaryotes (i.e., archaea and bacteria) [21] and eukaryotes [25]. The metabolic networks in these datasets were well identified and available from the KEGG database [23]. To prevent redundancies, when a species had different strains, we proceeded to use the strain whose genome was reported first as the representative strain for that species. Finally, 172 organisms, including 45 archaea, 60 bacteria, and 67 eukaryotes, were investigated (see Supplemental Material).

B. Construction of metabolic networks

The construction of metabolic networks follows similar protocols as those described in a previous study [22].

We downloaded XML files (version 0.7.1) containing the metabolic network data of 172 organisms on March 17, 2014 from the KEGG database [23]. On the basis of the metabolic information, substrate-product relationships were identified as carbon traces using the latest version [26] and an earlier version [2] of the metabolic reaction database. For comparison to previous studies (e.g., [11, 13]), the metabolic networks are represented by undirected networks (i.e., substrate graphs) in which the nodes and edges correspond to metabolites and reactions, respectively (i.e., substrate-product relationships are based on atomic mapping [3]). Currency metabolites such as H_2O , ATP, and NADH were excluded as described previously [27]. Moreover, the largest (weakly) connected component was extracted from each metabolic network to obtain more accurate calculations of $N_B(l_B)$ (i.e., to avoid bias from isolated components).

The constructed metabolic networks are available as a Supplemental Material.

C. Determination of network fractality

Network fractality is determined by $N_B(l_B)$ [11, 13, 28], calculated using a box-covering method. Covering a network with the minimum possible number of boxes is related to graph coloring [28], an NP-hard problem; thus, it requires approximation algorithms. Although Song et al. [28] have proposed several algorithms, we used the compact-box-burning (CBB) algorithm in this study because of the simple implementation of the CBB algorithm. $N_B(l_B)$ was averaged over 100 realizations. The selection of box covering algorithms poses little problem because all the algorithm in [28] can find the optimal solution with the similar accuracy.

The previous studies [11, 13, 28] have demonstrated that fractal and nonfractal networks show $N_B(l_B) \propto l_B^{-d_B}$ and $N_B(l_B) \propto \exp(-l_B)$ (because of a small-world property or logarithmic increase of the average diameter of a network with the total number of nodes N), respectively. We discriminate network fractality by comparing between the coefficients of determination, a measure of goodness-of-fit, of a power-law fit (R_{pow}^2) and exponential fit (R_{exp}^2) between $N_B(l_B)$ and l_B . R^2 is not influenced by sample size because it is an effect size in statistics.

R_{pow}^2 and R_{exp}^2 were calculated from the lm function in the R software (www.r-project.org) using the regression formulas $\ln N_B(l_B) = -d_B \ln l_B + b_1$ and $\ln N_B(l_B) = a_2 l_B + b_2$, respectively.

$R_{\text{pow}}^2 > R_{\text{exp}}^2$ and $R_{\text{pow}}^2 < R_{\text{exp}}^2$ indicate a fractal network and nonfractal one, respectively.

III. RESULTS

A. Metabolic networks are almost self-dissimilar

A comparison between R_{pow}^2 and R_{exp}^2 (Fig. 1 (a)) demonstrates that the latest version of metabolic networks are not fractal because $R_{\text{pow}}^2 < R_{\text{exp}}^2$ ($p = 2.2 \times 10^{-16}$ using the one-sample t -test) (see also Fig. 1 (b)). In particular, $N_B(l_B)$ is described by an exponential function rather than a power-law function (mean R_{exp}^2 is 0.97) (Fig. 2). As a representative example, the cases of *Escherichia coli*, yeast, and thale cress are shown. However, several exceptions were observed. For example, the latest metabolic network of thale cress, a plant, show the self-similarity.

The network representation in this study is slightly different from that in previous studies (e.g., [11, 13]). We defined edges as substrate-product pairs on the basis of carbon trances (see Sec II B) in this study. This approach was inspired by Arita's study [3], in which he pointed out that the pathways computed in the classical

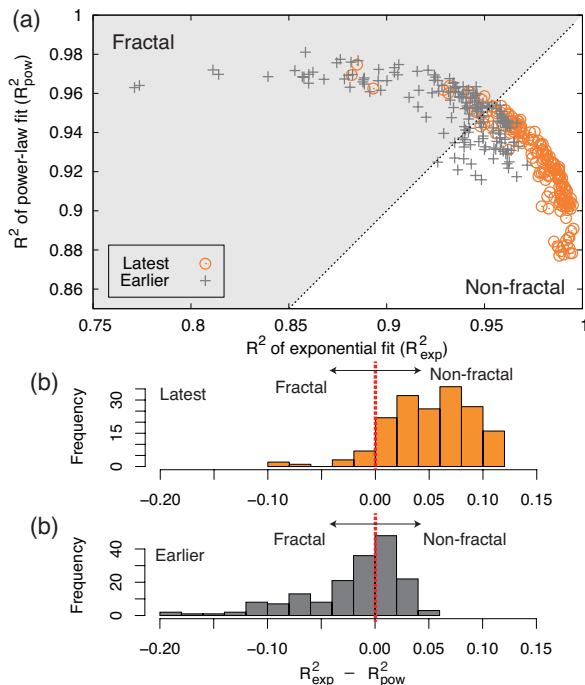


FIG. 1: (Color online) Comparison between the coefficients of determination of an exponential fit (R_{exp}^2) and power-law fit (R_{pow}^2). (a) Scatter plot of R_{exp}^2 versus R_{pow}^2 . Distributions of $R_{\text{exp}}^2 - R_{\text{pow}}^2$ in the earlier version (b) of latest version (c) of metabolic networks.

manners (i.e., network representations without consideration of atomic traces) do not conserve their structural moieties and, therefore, do not correspond to biochemical pathways on the traditional metabolic map. In addition to this, it remains possible that the definition of currency metabolites, such as water and ATP, is slightly different from that in previous studies (e.g., [11–14, 16, 17]). In particular, we considered the currency metabolites in our previous study [27], which are defined according to famous previous studies on metabolic networks (e.g., [1–3]). Since the previous studies used the metabolic network data in Ref. [1], it is expected that the definition is almost similar between our study and the previous studies. However, we could not conclude whether the definition is really similar because the definition has not been clearly described in the previous studies on network factuality.

To show that differences in network representation pose few problems, we also evaluated network factuality of the earlier version of the metabolic networks, using the box-counting method [28]. The earlier version of metabolic networks are fractal because $R_{\text{pow}}^2 > R_{\text{exp}}^2$ ($p = 1.1 \times 10^{-7}$ using the one-sample t -test) (see also Fig. 1 (c)), as reported in a number of previous studies (e.g., [11, 13]). Especially, $N_B(l_B)$ can be accurately described using a power-law function (mean R_{pow}^2 is 0.95) (Fig. 2). This result is consistent with the finding reported in the previous studies [11, 13]; thus, it indicates

that the procedures used for data analysis (e.g., the definition of currency metabolites) in this study were not problematic. On the other hand, the difference in analysis results between the earlier version and latest version of metabolic networks implies that the observed network fractality might result from a lack of data on metabolic reactions.

B. Differences in network measures between the earlier and latest versions of metabolic networks

We here discuss the effect of updates of the metabolic reaction database on network fractality.

Inspired by a relationship between disassortativity (i.e., negative degree–degree correlation) and network self-similarity [12], we first focused on the difference of an assortative coefficient r [29], defined as Pearson product-moment correlation coefficient of degree–degree correlations, between the earlier version (r_{earlier}) and latest version (r_{latest}) of metabolic networks: $\Delta r = r_{\text{latest}} - r_{\text{earlier}}$. However, such a difference could not be concluded ($p = 0.41$ using the one-sample t -test) although both versions of metabolic networks show a weak disassortativity (mean r of the earlier version and latest version of the networks are -0.011 ($p = 0.0082$ using the one-sample t -test) and -0.0080 ($p = 0.0054$ using the one-sample t -test), respectively). This result implies that the vanishment of network fractality cannot be explained in terms of assortative mixing in networks.

The database updates resulted in the metabolic networks of the latest version being larger than those of the earlier version. Thus, we next investigated a simpler network property: network density D , defined as E/N , where E and N are the number of nodes and the number of links, respectively.

We here considered the change ratio of a network measure X , defined as $R_X = X_{\text{latest}}/X_{\text{earlier}} - 1$, where X_{latest} and X_{earlier} are X obtained from the latest and earlier versions of metabolic networks, respectively. This definition had a limitation in that it did not consider the loss of nodes and edges because of the database update; however, this discrepancy was minor. The latest metabolic network included approximately 98% and 95% of nodes and edges, respectively, contained in the earlier version [22].

D_{latest} is significantly larger than D_{earlier} ($p < 2.2 \times 10^{-16}$ using the one-sample t -test). Furthermore, we found a negative correlation between R_D and ΔR_{pow}^2 , defined as $[R_{\text{pow}}^2]_{\text{latest}} - [R_{\text{pow}}^2]_{\text{earlier}}$, where $[R_{\text{pow}}^2]_{\text{latest}}$ and $[R_{\text{pow}}^2]_{\text{earlier}}$ are R_{pow}^2 obtained from the latest version and earlier version of metabolic networks, respectively. This result suggests that the increase in network density (i.e., database updates) decreases the self-similarity in metabolic networks.

Taken together, we speculate that the decrease in network fractality (i.e., R_{pow}^2) due to updating of the database occurred for the following reasons. In general,

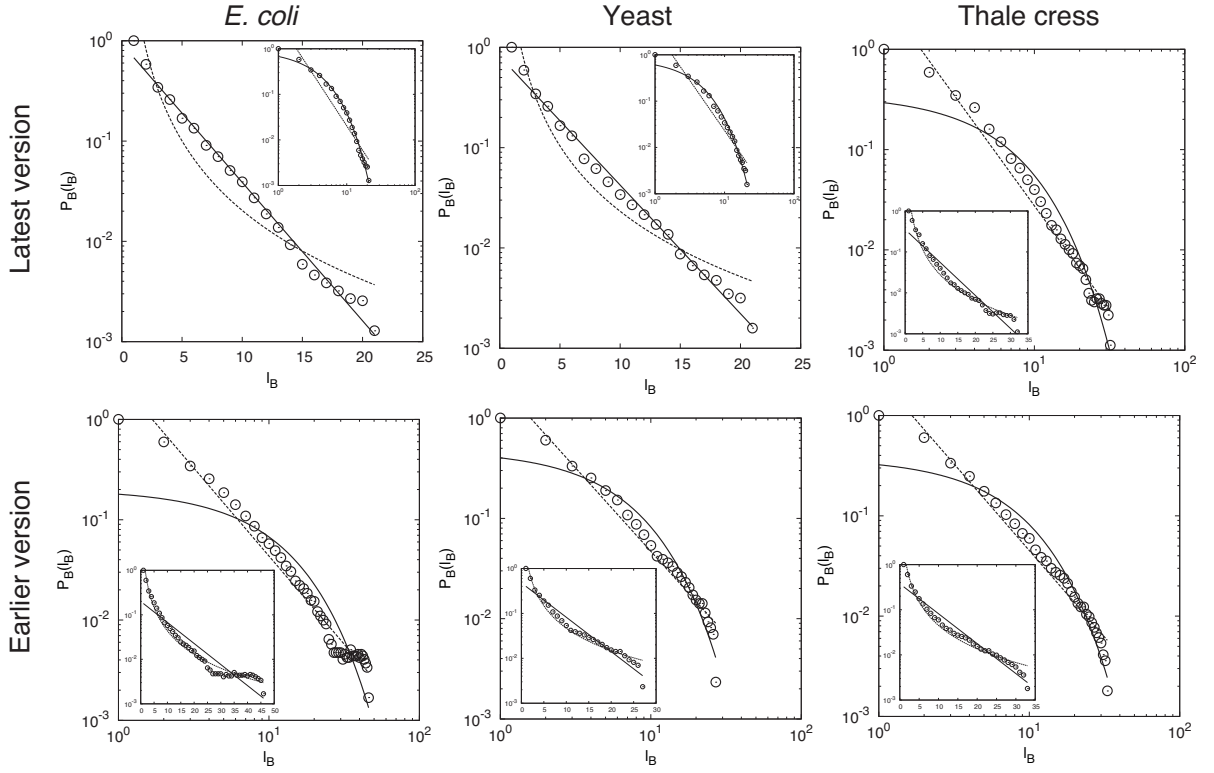


FIG. 2: Scatter plots of the normalized number of boxes (i.e., $P_B(l_B) = N_B(l_B)/N$) versus box size l_B in three representative organisms, obtained from the latest version (upper row) and earlier version (lower row) of metabolic networks: *Escherichia coli* (left column), *Saccharomyces cerevisiae* (yeast) (middle column), *Arabidopsis thaliana* (thale cress) (right column). Solid and dashed lines indicate exponential fits and power-law fits, respectively.

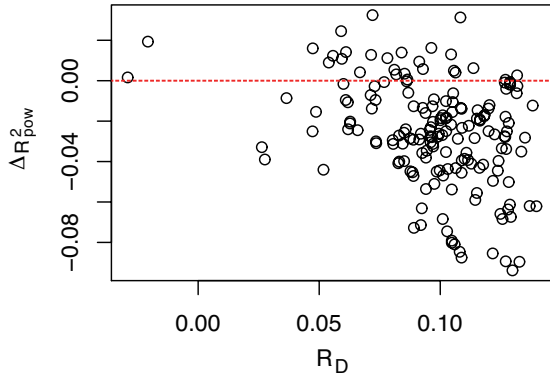


FIG. 3: (Color online) Negative correlation between ΔR_{pow}^2 and R_D (Spearman's rank correlation coefficient $r_s = -0.30$ and $p = 5.2 \times 10^{-5}$).

network fractality decreased slightly because metabolic networks may be randomized due to the addition of links among previously existing nodes. In this case, a number of structural properties are changed. In particular, Song et al. [13] have pointed out the importance of diameter in the origin of network factuality. Thus, an addition of links among boxes, detected in the earlier version of metabolic networks, decreases the network diameter and

average shortest path length; as a result, it collapses a fractal structure (i.e., it leads to an exponential decay of $N_B(l_B)$). This speculation may be supported by the observed positive correlations of ΔR_{pow}^2 with the absolute change ratio of the average shortest path length L (i.e., $|R_L|$) (Spearman's rank correlation coefficient $r_s = -0.36$ and $p = 1.5 \times 10^{-6}$) and absolute change ratio of the diameter d (i.e., $|R_d|$) ($r_s = -0.68$ and $p < 2.2 \times 10^{-16}$) (see Supplemental Material). Note that an absolute change ratio was used in this case because of these network measures have negative change ratios.

C. A part of eukaryotic metabolic networks are fractal, but it is still debatable

Although the latest version of metabolic networks tends to be nonfractal, there are several exceptions (Figs. 1 and 2). To reveal a tendency of these exceptions, we investigated metabolic network fractality in grater details, according to species classifications.

We first focused on three domains of life: archaea, bacteria, and eukaryotes (Fig. 4). The archaeal and bacterial (i.e., prokaryotic) metabolic networks are significantly self-dissimilar because $R_{\text{exp}}^2 > R_{\text{pow}}^2$ in both cases of archaea and bacteria ($p = 1.6 \times 10^{-11}$ and

$p < 2.2 \times 10^{-16}$ using the one-sample t -test, respectively). Overall, the eukaryotic metabolic networks are also non-fractal because $R_{\text{exp}}^2 > R_{\text{pow}}^2$ ($p = 2.2 \times 10^{-9}$); however, the distribution of $R_{\text{exp}}^2 - R_{\text{pow}}^2$ in the eukaryotic networks is broader than that in the prokaryotic ones and more exceptions are observed in the eukaryotic case. A degree of metabolic network change due to database updates may be not able to explain this difference in the distributions because the mean change ratio R_D in prokaryotes (archaea and bacteria) (0.099) is almost equivalent to that in eukaryotes (0.096) ($p = 0.52$ using the two-sample t -test). That is, it is difficult to conclude that eukaryotic metabolic networks are relatively self-similar because they are relatively unchanged despite database updates.

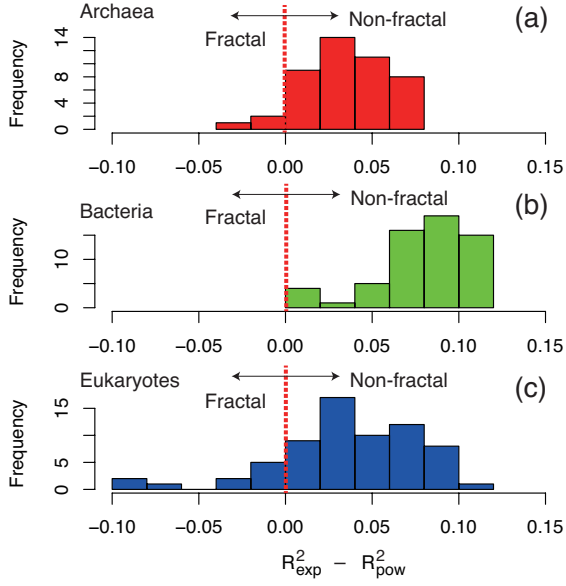


FIG. 4: (Color online) Distributions of $R_{\text{exp}}^2 - R_{\text{pow}}^2$, obtained from the latest version of metabolic networks, in archaea (a), bacteria (b), and eukaryotes (c).

Thus, we next investigated on eukaryotic metabolic network fractality in greater details, according to a eukaryotic classification (Fig. 5), defined in the KEGG database (i.e., plants, animals, fungi, and protists), and found that plant metabolic networks show fractality although the following plant species are exceptions: *Ostreococcus lucimarinus* (olu), *Chlamydomonas reinhardtii* (cre), and *Cyanidioschyzon merolae* (cme). Since these plant species are microorganisms, they can be distinguished from the other plants, which are higher plants such as trees and flowers. Thus, it concluded that metabolic networks of higher plants are fractal because $R_{\text{pow}}^2 > R_{\text{exp}}^2$ ($p = 0.034$ using the one-sample t -test). However, a degree of metabolic network change due to database updates (i.e., mean R_D) is different between the higher plants (0.081) and the other eukaryotes (0.098) ($p = 0.010$ using the two-sample t -test); thus, it remains possible that metabolic networks of higher plants are

still self-similar because they show less change despite database updates.

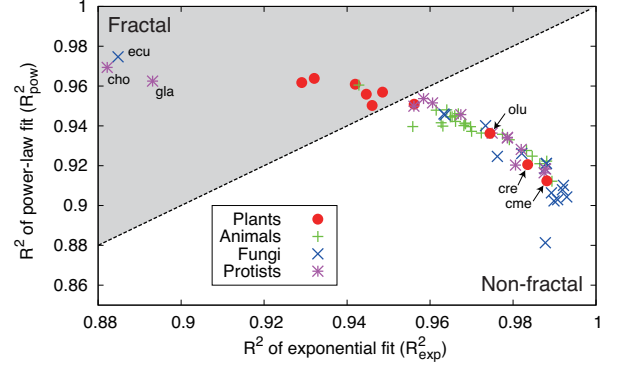


FIG. 5: (Color online) Scatter plot of R_{exp}^2 versus R_{pow}^2 in the latest version of eukaryotic metabolic networks.

On the other hand, metabolic networks are self-dissimilar (i.e., $R_{\text{exp}}^2 > R_{\text{pow}}^2$) in animals and fungi ($p = 1.2 \times 10^{-7}$ and $p = 4.8 \times 10^{-6}$ using the one-sample t -test, respectively); however, we cannot determine whether protist metabolic networks are fractal or nonfractal because $R_{\text{pow}}^2 = R_{\text{exp}}^2$ ($p = 0.55$ using the one-sample t -test). In particular, the metabolic networks of *Encephalitozoon cuniculi* (ecu), *Cryptosporidium hominis* (cho), and *Giardia lamblia* (gla) show the self-similarity, exceptionally. However, the metabolic networks of these eukaryotes were almost unchanged due to database updates. Especially, mean R_D of these eukaryotes is 0.0057, and it is significantly smaller than that of other eukaryotes (0.098).

Taken together, the metabolic networks of eukaryotes are also self-dissimilar. Although metabolic network fractality can be concluded in a part of eukaryotes (higher plants, in particular); however, it is still debatable.

IV. DISCUSSION

In this study, we showed that metabolic networks are almost self-dissimilar, contrary to a number of previous studies [11–17]. Rather, this result strongly supports the scale-richness and self-dissimilarity in metabolic networks, proposed by Tanaka [19, 20]. The definition of modules or boxes in Tanaka's studies [19, 20] is arbitrary (although it is based on biological knowledge), and it is different from that in previous studies (e.g., [11, 13]), which used box-covering methods. Thus, it is difficult to make an easy comparison between these previous studies. In addition to this, these previous studies only focused on a limited number of species. Our comprehensive evaluation provided a more convincing evidence of the self-dissimilarity of metabolic networks by avoiding these limitations in these previous studies.

We demonstrated that the previously observed network fractality was probably due to a lack of available

data on metabolic reactions. Especially, metabolic networks have not been fully understood; thus, there is a need for a more careful examination in data analysis in the future. For example, enzyme promiscuity [30], which implies that enzymes can catalyze multiple reactions, act on more than one substrate, or exert a range of suppressions [31], in which an enzymatic function is suppressed by overexpressing enzymes showing originally different functions, suggests the existence of many hidden metabolic reactions. Consideration of these hidden metabolic reactions is important for understanding metabolic network fractality. In addition to this, these experimental studies suggest that metabolic networks are more flexible than previously thought. Moreover, previous theoretical and data analytic studies [21, 32–34] argue that specific features in the networks are weakly correlated with a system-specific purpose, function, or causal chain. These results imply that metabolic networks more randomly constructed than previously thought, and they confirm our finding that an increase of network density enhances network randomization (Fig. 3).

In addition to this, the earlier version of metabolic networks show self-similarity although it shows a very weak disassortativity, and the network fractality vanished due to database updates although the assortative coefficient was unchanged (Sec. IIIB), this result implies a limitation of disassortativity as determination factor of network self-similarity [12], and it underscores the need for a more careful discussion of the origin of self-similarity of complex networks. Especially, it may suggest the existence of different possible origins of network fractality.

This study does not discount the importance of self-similarity of metabolic networks. Rather, it emphasizes the need for a more suitable definition of network fractality and a more careful examination of self-similarity of metabolic networks. For example, this study does not consider several important properties of metabolic networks, as do many other works on metabolic network analyses: reaction stoichiometry, the direction of reaction (i.e., reversible versus irreversible), chemical structure of metabolites, and gene clusters. It will be important to propose box-covering (renormalization) methods that also consider such kinds of biological information in metabolic networks. In this context, methods for finding modular architecture (i.e., biologically meaningful boxes) of metabolic pathways based on gene clusters and chemical transformation patterns [35, 36] may be useful. Using these definitions, we may be able to evaluate biologically understandable fractality in metabolic networks and other biological networks.

Acknowledgments

This study was supported by a Grant-in-Aid for Young Scientists (A) from the Japan Society for the Promotion of Science (Grant No. 25700030). K.T. was partly supported by Chinese Academy of Sciences Fellowships for Young International Scientists (Grant No. 2012Y1SB0014), and the International Young Scientists Program of the National Natural Science Foundation of China (Grant No. 11250110508).

-
- [1] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, *Nature* **407**, 651 (2000).
 - [2] H. Ma and A. P. Zeng, *Bioinformatics* **19**, 270 (2003).
 - [3] M. Arita, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1543 (2004).
 - [4] A.-L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).
 - [5] R. Albert, *J. Cell. Sci.* **118** 4947 (2005).
 - [6] K. Takemoto, *Metabolites* **2**, 429 (2012).
 - [7] W. Winterbach, P. van Mieghem, M. Reinders, H. Wang, and D. de Ridder, *BMC Syst. Biol.* **7**, 90 (2013).
 - [8] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [9] E. Ravasz and A.-L. Barabási, *Phys. Rev. E* **67**, 026112 (2003).
 - [10] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A.-L. Barabási, *Science* **297**, 1551 (2002).
 - [11] C. Song, S. Havlin, and H. A. Makse, *Nature* **433**, 392 (2005).
 - [12] S.-H. Yook, F. Radicchi, and H. Meyer-Ortmanns, *Phys. Rev. E* **72**, 045105 (2005).
 - [13] C. Song, S. Havlin, and H. A. Makse, *Nat. Phys.* **2**, 275 (2006).
 - [14] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **96**, 018701 (2006).
 - [15] L. K. Gallos, C. Song, and H. A. Makse, *Physica A* **386**, 686 (2007).
 - [16] L. K. Gallos, C. Song, S. Havlin, and H. A. Makse, *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7746 (2007).
 - [17] Y. Sun and Y. Zhao, *Phys. Rev. E* **89**, 042809 (2014).
 - [18] M. Arita, *J. Biochem.* **138**, 1 (2005).
 - [19] R. Tanaka and J. Doyle, *arXiv: q-bio/0410009*.
 - [20] R. Tanaka, *Phys. Rev. Lett.* **94**, 168101 (2005).
 - [21] K. Takemoto, *Phys. Rev. E* **86**, 036107 (2012).
 - [22] K. Takemoto, *PLoS One* **8**, e61348 (2013).
 - [23] M. Kanehisa, et al., *Nucleic Acids Res.* **42**, D199 (2014).
 - [24] R. Caspi, et al., *Nucleic Acids Res.* **40**, 742 (2012).
 - [25] K. Takemoto and I. Yoshitake, *Metabolites* **3**, 979 (2013).
 - [26] M. Stelzer, J. Sun, T. Kamphans, S. P. Fekete, and A.-P. Zeng, *Integr. Biol.* **3**, 1071 (2011).
 - [27] K. Takemoto, J. C. Nacher, and T. Akutsu, *BMC Bioinformatics* **8**, 303 (2007).
 - [28] C. Song, L. K. Gallos, S. Havlin, and H. A. Makse, *J. Stat. Mech.* **2007**, P03006 (2007).
 - [29] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
 - [30] O. Khersonsky and D. S. Tawfik, *Annu. Rev. Biochem.* **79**, 471 (2010).
 - [31] W. M. Patrick, E. M. Quandt, D. B. Swartzlander, and I. Matsumura, *Mol. Biol. Evol.* **24**, 2716 (2007).
 - [32] P. Minnhagen and S. Bernhardsson, *PLoS One* **3**, e1690

- (2008).
- [33] S. Bernhardsson and P. Minnhagen, New J. Phys. **12**, 103047 (2010).
- [34] S. H. Lee, S. Bernhardsson, P. Holme, B. J. Kim, and P. Minnhagen, New J. Phys. **14**, 033032 (2012).
- [35] M. Kanehisa, FEBS Lett. **587**, 2731 (2013).
- [36] A. Muto, et al., J. Chem. Inf. Model. **53**, 613 (2013).